Astronomy & Astrophysics manuscript no. 20937
April 1, 2013

Crucial aspects of the initial mass function (II)

The inference of total quantities from partial information on a cluster

Miguel Cerviño[1,2], Carlos Román-Zúñiga[3], Amelia Bayo[4,5], Valentina Luridiana[2,6], Néstor Sánchez[7] and Enrique Pérez[1]

[1] Instituto de Astrofísica de Andalucía (IAA-CSIC), Glorieta de la Astronomía s/n, 18008 Granada, Spain
[2] Instituto de Astrofísica de Canarias, c/ vía Láctea s/n, 38205 La Laguna, Tenerife, Spain
[3] Instituto de Astronomía, Universidad Académica en Ensenada, Universidad Nacional Autónoma de México, Ensenada BC, 22860 Mexico
[4] European Southern Observatory, Casilla 19001, Santiago 19, Chile
[5] Max Planck Institut für Astronomie, Königstuhl 17, 69117, Heidelberg, Germany
[6] Departamento de Astrofísica, Universidad de La Laguna (ULL), 38205 La Laguna, Tenerife, Spain
[7] S. D. Astronomía y Geodesia, Fac. CC. Matemáticas, Universidad Complutense de Madrid, 28040, Madrid, Spain.

Received ; accepted

ABSTRACT

Context. In a probabilistic framework of the interpretation of the initial mass function (IMF), the IMF cannot be arbitrarily normalized to the total mass, $\mathcal{M}$, or number of stars, $\mathcal{N}$, of the system. Hence, the inference of $\mathcal{M}$ and $\mathcal{N}$ when partial information about the studied system is available must be revised. (i.e., the contribution to the total quantity cannot be obtained by simple algebraic manipulations of the IMF).
Aims. We study how to include constraints in the IMF to make inferences about different quantities characterizing stellar systems. It is expected that including any particular piece of information about a system would constrain the range of possible solutions. However, different pieces of information might be irrelevant depending on the quantity to be inferred. In this work we want to characterize the relevance of the priors in the possible inferences.
Methods. Assuming that the IMF is a probability distribution function, we derive the sampling distributions of $\mathcal{M}$ and $\mathcal{N}$ of the system constrained to different types of information available.
Results. We show that the value of $\mathcal{M}$ that would be inferred must be described as a probability distribution $\Phi_{\mathcal{M}}[\mathcal{M}; m_{\rm a}, N_{\rm a}, \Phi_{\mathcal{N}}(\mathcal{N})]$ that depends on the completeness limit of the data, $m_{\rm a}$, the number of stars observed down to this limit, $N_{\rm a}$, and the prior hypothesis made on the distribution of the total number of stars in clusters, $\Phi_{\mathcal{N}}(\mathcal{N})$.

Key words. stars: statistics — galaxies: stellar content — methods: data analysis

1. Introduction

The study of cluster dynamics and star formation relies on the knowledge of cluster masses and the amount of such mass transformed into stars, $\mathcal{M}$. In most cases, we have partial information of the system, i.e., the observations of some stars in the cluster. Such information is usually used in the inverse problem using the initial mass function (IMF) *realization* (see below) as a distribution by number to make inferences about a theoretical probability distribution function, the IMF $\phi(m)$ (Bouvier et al. 1998; Briceño et al. 2002; Luhman et al. 2003; Oliveira et al. 2009; Bayo et al. 2011). However, such information is not enough to obtain cluster masses, and for some astrophysical studies it is required to assume a $\phi(m)$ covering all the range of possible stellar masses to make inferences about global cluster properties (the direct problem).

This use of the term IMF for both the distribution by number for the inverse problem of statistics and the probability distribution function (pdf) for the direct problem can lead to different interpretations of the IMF itself and the results obtained from it (cf. Cerviño et al. 2012, hereafter Paper I). In this work, following Scalo (1986), we will adopt the pdf definition[1].

The shape of the pdf and that of the distribution by number depend crucially on the size of the sample, that is, the number of stars $\mathcal{N}$; for large $\mathcal{N}$ values, the two shapes tend to be similar. However, this similarity can mislead one into believing that the distribution by number is just a scaled-up version of the pdf, with $\mathcal{N}$ being the scale factor. This would be very wrong since the physical meanings of both distributions are intrinsically different; Paper I is dedicated to exploring the consequences of this essential difference.

As a consequence, the standard methodology used to infer $\mathcal{M}$ values, which assumes the use of a correction factor for unobserved stars, is no longer valid. The main goal of this paper is to define a methodology based on the probabilistic approach of the IMF to obtain the total stellar mass $\mathcal{M}$ of an stellar sample from limited information on the sample itself.

This task is far from trivial as we have to bridge different gaps according to the amount of unknown information. We start the discussion by making an inventory of possible scenarios that differ from each other according to the amount of information

Send offprint requests to: M. Cerviño e-mail: `mcs@iaa.es`

[1] This definition implies that stellar masses are identically and independent distributed, we refer Paper I for more details.

available, with the aim of emphasizing how this affects the determination of $\mathcal{M}$ and $\mathcal{N}$. Five such scenarios are:

1. We know (from the IMF) the probability of a random star having a mass $m_{\rm star}$ equal to or larger than some given value m_a, but no specific information on the particular cluster is known.
2. We know (from observations) the number of stars $\mathcal{N}$ in a particular cluster; we also know (from the IMF) the expected number of stars with $m \geq m_{\rm a}$.
3. We know (from observations) the number of stars $\mathcal{N}$ in a particular cluster; we also know (from observations, too) that $N_{\rm a}$ stars have $m \geq m_{\rm a}$ and the mass of such stars.
4. We know that a particular cluster has $N_{\rm a}$ stars with $m \geq m_{\rm a}$ and the mass of such stars from observations.
5. We know that a particular cluster has $N_{\rm a}$ stars with $m \geq m_{\rm a}$ and the mass of such stars, and we also know its total mass $\mathcal{M}$.

In scenario 1, which relies solely on knowledge of the IMF, we only know a theoretical probability that is independent of $\mathcal{N}$ and $\mathcal{M}$. Consecuently, we have neither information on $\mathcal{M}$ nor on the actual value of $m_{\rm star}$.

In scenario 2, we know that the cluster is the result of sampling the IMF with $\mathcal{N}$ stars. With such information, we can compute the sampling distribution of $\mathcal{M}$: that is, the distribution of possible values of $\mathcal{M}$ constrained by the value of $\mathcal{N}$. In particular, if $\mathcal{N} = 1$ the distribution of total masses is the IMF itself, and if $\mathcal{N} \to \infty$, the distribution of $\mathcal{M}$ is a Gaussian, because of the central limit theorem. In all intermediate cases, the sampling distribution of $\mathcal{M}$ at a given $\mathcal{N}$ is a more or less asymmetric function, which in turn implies that its mean value $\langle\mathcal{M}\rangle$ is not (in general) the same as its most probable value.

Scenario 3 is a constrained version of the previous one. In the universe of all possible clusters with $\mathcal{N}$ stars, only those *conditioned to* have $N_{\rm a}$ stars with mass equal to or larger than $m_{\rm a}$ can represent the cluster studied. The resulting distribution of possible $\mathcal{M}$, which is different from the previous sampling distribution, can be obtained by imposing an a posteriori condition on it. However, since we also know the mass of the $N_{\rm a}$ stars, an additional constraint must be applied

Scenario 4 only constrains $\mathcal{M}$ to be equal to or larger than the contribution of the $N_{\rm a}$ stars. We cannot progress further unless we additionally assume a distribution of possible $\mathcal{N}$ values. If we do so, the resulting values of the mean total mass $\langle\mathcal{M}\rangle$ and the most probable value will differ from those obtained under scenario 2, since in the present case $\mathcal{N}$ is not fixed but distributed and this affects the shape of the sampling distribution of $\mathcal{M}$.

In scenario 5, we know that the mass is $\mathcal{M}$ and that there are $N_{\rm a}$ stars with $m \geq m_{\rm a}$. The probability distributions that describe such a cluster (such as, for example, the distribution of possible $\mathcal{N}$ values or of the $N_{\rm a}$ most massive stars that the cluster could host) correspond to the particular situation described in scenario 4 with the additional constraint of knowing $\mathcal{M}$.

From the above discussion, it is clear that the $\mathcal{M}$ derived in each of the above scenarios are different. Although all the resulting distributions are derived from the IMF, each of them is the result of including different pieces of information in the analysis: either the total number of stars $\mathcal{N}$ in the cluster (in scenarios 2 and 3), and its probability distribution (in scenarios 4 and 5) or the presence of $N_{\rm a}$ stars above a given mass value (in scenarios 3, 4, and 5). Each case results in a different conditional probability distribution, which results in a different estimation of $\mathcal{M}$.

We note that relating the IMF with the corresponding sampling (and conditional) probability distributions is correct, given the set under study. We also note that we have an additional piece of information in such a set: stars are individual, discrete entities (i.e., $\mathcal{N}$ is a natural number). Such a condition must be fulfilled by any cluster in the Universe and must be included in all scenarios as a restriction (even in cases where there is no explicit reference to $\mathcal{N}$, as in scenario 5).

The preceding discussion boils down to the following point: as an underlying density distribution, the IMF describes neither a particular case nor any observational constraints (such as, e.g., the number of stars with a given mass observed in a particular cluster). Once an observational constraint is included (e.g., the fact that one star with known mass is present), conditional probabilities must be applied. Stated otherwise, the distribution that describes the universe of possible results (the IMF) is an a priori probability, and the probability constrained to the observed data is an a posteriori (conditional) probability. Confusing the a posteriori probability with the a priori probability is one of the most common flaws in hypothesis testing reasoning (this is also called the *Prosecutor's fallacy*: see Selman & Melnick 2008, for a discussion in a similar astrophysical context). In these situations, it is fundamental understand the true context of the question before seeking an answer. This has been done in the five scenarios discussed above.

The structure of the paper is as follows: In Sect. 2 we summarize the basic concepts required to use the IMF in a probabilistic framework (see Paper I for a more extended discussion). In Sect. 3 we consider an ideal case in which all the stars in the system are known. Then we replace known information by unknowns to describe real situations where the use of the IMF or a related sampling distribution is required. Section 4 shows the methodology to obtain $\mathcal{M}$ from partial information of the system in the scenarios presented above and their application to some astrophysical cases. We discuss some considerations about the use of prior information in Sect. 5. Our conclusions are described in Sect. 6.

2. Formal probabilistic formulation

The basis of the probabilistic formulation has been presented in Paper I. We refer to that paper for more details and include here only the basic formulae needed for this work.

1. The IMF, $\phi(m) = {\rm d}N/{\rm d}m$, is a probability density function (pdf), which can be integrated over a given mass range to derive the probability of finding a star in that range. The mass limits $\rm m_{low}$ and $\rm m_{up}$ are given by stellar theory and must fulfill $\int_{\rm m_{low}}^{\rm m_{up}} \phi(m){\rm d}m = 1$; that is, we are certain that any possible star has a mass between $\rm m_{low}$ and $\rm m_{up}$.
   The probability of a random star having a mass *lower than* a given value $m_{\rm a}$ is given by

$$p(m < m_{\rm a}) = \int_{\rm m_{low}}^{m_{\rm a}} \phi(m)\,{\rm d}m. \quad (1)$$

   In this work, the integrals over the IMF will always be read as *equal to or larger than* the lower limit and *lower than* the upper limit.
   In this work we employ the Kroupa IMF (Kroupa 2001, 2002) as used in Weidner & Kroupa (2006), with $\rm m_{up}$ =

$120\mathrm{M}_\odot$, $\mathrm{m}_{\mathrm{low}} = 0.01\mathrm{M}_\odot$, and a correction of $k' = 1/3$ for stars with mass lower than $0.08\ \mathrm{M}_\odot$[2]

2. Different observational scenarios can be described by adding constraints to the IMF. For instance, we may explicitly include a limit on m_a and compute probabilities for stars with masses lower than m_a. In this case, we must define an a posteriori pdf related to the IMF that includes such a condition:

$$\phi(m|m < m_\mathrm{a}) = \frac{\phi(m)\,\mathrm{H}(m_\mathrm{a} - m)}{p(m < m_\mathrm{a})}, \tag{2}$$

where $\mathrm{H}(m_\mathrm{a} - m)$ is the Heaviside function[3], which ensures that no star equal to or larger than m_a can be present in the cluster. We note that $\phi(m|m < m_\mathrm{a})$ is a pdf also. The mean mass of such distribution is

$$\langle m|m < m_\mathrm{a}\rangle = \frac{\int_{\mathrm{m_{low}}}^{\mathrm{m_{up}}} m\,\phi(m)\,\mathrm{H}(m_\mathrm{a} - m)\mathrm{d}m}{p(m < m_\mathrm{a})}. \tag{3}$$

3. The pdf describing ensembles with a total number of stars $\mathcal{N}$ (formally, a sampling distribution conditioned to have $\mathcal{N}$ stars) can be calculated as successive convolutions of the corresponding pdf for one star. For instance, the pdf for the total mass, $\Phi_{\mathcal{M}}(\mathcal{M}|\mathcal{N})$, is the result of convolving the IMF $\mathcal{N}$ times in a recursive convolution (see Cerviño & Luridiana 2006; Selman & Melnick 2008):

$$\Phi_{\mathcal{M}}(\mathcal{M}|\mathcal{N}) = \overbrace{\phi(m) \otimes \phi(m) \otimes\ ....\ \otimes \phi(m)}^{\mathcal{N}}. \tag{4}$$

The same procedure applies to any other pdf. The mean value of the resulting distribution is

$$\langle \mathcal{M}|\mathcal{N}\rangle = \mathcal{N} \times \langle m\rangle = \mathcal{N} \times \int_{\mathrm{m_{low}}}^{\mathrm{m_{up}}} m\,\phi(m)\,\mathrm{d}m. \tag{5}$$

Mean values of constrained distributions when sampled with $\mathcal{N}$ stars are obtained in a similar way.

3. The trade-off between knowledge and probability

Once we have laid down the basic framework, we apply it to our science case: the estimation of the total mass $\mathcal{M}$ of a cluster from a partial knowledge of its stellar content. To do that we progressively replace known information by unknowns to describe real situations; however, the following items here are not directly related to the scenarios quoted in the Introduction (we will come back to such scenarios in Sect. 4).

3.1. Case study 1: Everything is known

We begin with an ideal observational point of view, where we suppose that we know the masses m_i^{obs} of every one of the $\mathcal{N}$ stars in a cluster. Thus, the total mass, $\mathcal{M}$, is also known. In this hypothetical case, it is not required to use the IMF. However, this exercise allows us to illustrate the trade-off between the use of known data from a particular cluster (i.e., a particular IMF realization) and the use of probability distributions.

We sort the stars in ascending order according to their mass. We use a subindex in brackets to denote that such operation has been performed, so m_i is the i-th random sampled element and $m_{[i]}$ is the i-th element after sorting the data. We also assume that the most massive star has a mass $m_{[\mathcal{N}]}^{\mathrm{obs}} = m_{\mathrm{max}}^{\mathrm{obs}}$ with a value lower than $\mathrm{m_{up}}$.

In addition, we assume that we have N_a stars equal or more massive than an arbitrary value m_a, so that $m_{[\mathcal{N}-N_\mathrm{a}]} < m_\mathrm{a}$, and $m_{[\mathcal{N}-N_\mathrm{a}+1]} \geq m_\mathrm{a}$. We express the total number of stars and total mass as a function of the N_a set. It can be described as

$$N_\mathrm{a} = \sum_{i=\mathcal{N}-N_\mathrm{a}+1}^{\mathcal{N}} \delta_{i,i}, \quad M_\mathrm{a} = \sum_{i=\mathcal{N}-N_\mathrm{a}+1}^{\mathcal{N}} m_{[i]}^{\mathrm{obs}}\delta_{i,i}, \tag{6}$$

where $\delta_{i,j}$ is the Kronecker delta. The total mass in the ensemble is

$$\mathcal{M} = M_\mathrm{a} + \sum_{i=1}^{\mathcal{N}-N_\mathrm{a}} m_{[i]}^{\mathrm{obs}}\delta_{i,i}. \tag{7}$$

These two equations, rewritten in terms of frequencies and mean stellar mass *in the complete sample*, are, respectively

$$\frac{N_\mathrm{a}}{\mathcal{N}} = \sum_{i=\mathcal{N}-N_\mathrm{a}+1}^{\mathcal{N}} \frac{\delta_{i,i}}{\mathcal{N}}, \tag{8}$$

and

$$\langle\widetilde{m}\rangle = \frac{N_\mathrm{a}}{\mathcal{N}}\,\frac{M_\mathrm{a}}{N_\mathrm{a}} + \frac{\mathcal{N} - N_\mathrm{a}}{\mathcal{N}} \sum_{i=1}^{\mathcal{N}-N_\mathrm{a}} \frac{m_{[i]}^{\mathrm{obs}}}{\mathcal{N} - N_\mathrm{a}}. \tag{9}$$

Multiplying $\langle\widetilde{m}\rangle$ by $\mathcal{N}$ produces the value of $\mathcal{M}$. However, we note that conceptually

$$\mathcal{M} = \mathcal{N} \times \langle\widetilde{m}\rangle \neq \mathcal{N} \times \langle m\rangle = \langle\mathcal{M}\rangle, \tag{10}$$

since $\langle\widetilde{m}\rangle$ (the sample mean) does not coincide with the the mean stellar mass obtained from the IMF, $\langle m\rangle$ (the population mean). That is, $\langle\widetilde{m}\rangle$ is an estimate of $\langle m\rangle$ obtained from a sample of $\mathcal{N}$ stars, so, formally, $\langle\widetilde{m}\rangle = \langle\widetilde{m}|\mathcal{N}\rangle$. In the following, we use the $\widetilde{m}$ symbol to denote an estimate of m. In the computation of this estimate, the value of $\mathcal{N}$ must be taken into consideration, although we will not write it explicitly in order to simplify the notation.

3.2. Case study 2: The total number of stars and the mass of the most massive N_a stars are known

In this case we have less information than in the previous case since we only know $m_{[i]}^{\mathrm{obs}}$ with $i = \{\mathcal{N} - N_\mathrm{a} + 1, \ldots \mathcal{N}\}$, stellar masses, and $\mathcal{N}$. But we had seen that estimates obtained from actual values, such as $\langle\widetilde{m}\rangle$ can be related to values obtained from the IMF. So we can replace these estimates with

$$\sum_{i=1}^{\mathcal{N}-N_\mathrm{a}} \frac{m_i}{\mathcal{N} - N_\mathrm{a}} = \left\langle \widetilde{m}|m < m_\mathrm{a}^{\mathrm{obs}}\right\rangle \to \left\langle m|m < m_\mathrm{a}^{\mathrm{obs}}\right\rangle.$$

[2] Such a correction was not used in Paper I. However, it is the parametrization used in the set of clusters by Kirk & Myers (2011) we use in this work for comparing methodologies.

[3] We use here the Heaviside function as a distribution to define the domain of $\phi(m)$ including constraints. In this situation the value of H(0) is *not* defined, but it is assigned a posteriori to be consistent with the convention used in the integral limits. In the case of Eq. 2, H(0) = 0.

Thus, although we cannot know the actual $\mathcal{M}$ value, we can at least obtain average values given different sets of constraints:

$$\left\langle \mathcal{M} | m_{[i]}^{\rm obs} \geq m_{\rm a}^{\rm obs}\ i = \mathcal{N} - N_{\rm a} + 1, ...\mathcal{N};\ \mathcal{N} \right\rangle = M_{\rm a} + (\mathcal{N} - N_{\rm a}) \left\langle m | m < m_{\rm a}^{\rm obs} \right\rangle . \quad (11)$$

This illustrates the trade-off between observed frequency distributions and probability: when we use a probability distribution, we cannot have access to the actual values, but we can have access to the distribution of *possible* values and the mean value of *all* these possible values. In this case we are using the estimates argument in the opposite direction to a statistical analysis, i.e., we are making the assumption that *all* the stars are distributed following the IMF[4] and using it to make inferences about related quantities.

3.3. Case study 3: Only the mass of the $N_{\rm a}$ more massive stars is known

Observations of clusters in many cases only allow characterization of the $N_{\rm a}$ most luminous stars with masses $m_{[i]}^{\rm obs}$, $i = \{\mathcal{N} - N_{\rm a} + 1, \ldots \mathcal{N}\}$. They also lack a proper census that includes the lowest luminous members (see Bayo et al. 2011; Kirk & Myers 2011, as counterexamples). In this case, it is more difficult to obtain estimates, since we can not define a *frequency* of $N_{\rm a}$. Therefore, is the following reasoning valid?

$$\widetilde{p}(m|m < m_{\rm a}^{\rm obs}) = \frac{\mathcal{N} - N_{\rm a}}{\mathcal{N}} \rightarrow p(m|m < m_{\rm a}^{\rm obs}),$$
$$\widetilde{p}(m|m \geq m_{\rm a}^{\rm obs}) = \frac{N_{\rm a}}{\mathcal{N}} \rightarrow p(m|m \geq m_{\rm a}^{\rm obs}).$$

3.3.1. When is the correspondence $\mathcal{N} = N_{\rm a}/p(m|m \geq m_{\rm a})$ valid?

We divide the IMF in, e.g., $k+1$ mass intervals, where the mass interval containing the lower masses, e.g., the k+1, comprises the $\mathcal{N} - N_{\rm a}$ of unknown stars with mass lower than $m_{\rm a}$. Each of the remaining i mass interval, which belong to $[m_i^{\rm low}, m_i^{\rm up})$ contain n_i stars[5], so that $\sum_{i=1}^{k} n_i = N_{\rm a}$. The probability of having a star in a given mass interval is given by the integration of the IMF over such a mass interval, $p_i(m) = p(m \in [m_i^{\rm low}, m_i^{\rm up}))$. We assume that the cluster is a random realization of the IMF for $\mathcal{N}$ stars, so the probability of having the $\mathcal{N}$ stars distributed in the $k+1$ intervals with n_i stars in the i-th interval for a given (unknown) number of stars $\mathcal{N}$ is given by the multinomial distribution[6]

$$\Phi_{N_{\rm a}}(N_{\rm a}|\mathcal{N}) = \mathcal{P}(m \geq m_{\rm a}, \sum_{i=1}^{k} n_i = N_{\rm a}|\mathcal{N}) = \frac{\mathcal{N}!}{(\mathcal{N} - N_{\rm a})!\prod_{i=1}^{k} n_i!} p(m < m_{\rm a})^{\mathcal{N}-N_{\rm a}} \prod_{i=1}^{k} p_i(m)^{n_i}$$

$$= A(p_i, n_i)\ \frac{\mathcal{N}!}{(\mathcal{N} - N_{\rm a})!} p(m < m_{\rm a})^{\mathcal{N}-N_{\rm a}}, \quad (12)$$

where we have included in $A(p_i, n_i)$ all the known information. However, we are interested in the complementary distribution $\Phi_{\mathcal{N}}(\mathcal{N}|N_{\rm a})$, which must be obtained using the Bayes' theorem (see, e.g., Paper I). Assuming that the possible values of $\mathcal{N}$, $\Phi_{\mathcal{N}}(\mathcal{N})$ follow a discrete power-law probability distribution with exponent $-\beta$, we obtain

$$\Phi_{\mathcal{N}}(\mathcal{N}|N_{\rm a}) = A'\ \frac{\mathcal{N}!}{(\mathcal{N} - N_{\rm a})!}\ p(m < m_{\rm a})^{\mathcal{N}-N_{\rm a}}\ \mathcal{N}^{-\beta}, \quad (13)$$

where A' is a normalization value that includes all the known terms and is independent of $A(p_i, n_i)$ since $A(p_i, n_i)$ is canceled out by the normalization constant, Thus, the inference about the total number of stars only depends on the number of stars $N_{\rm a}$ more massive than a certain observational value $m_{\rm a}$, and not on the particular distribution of such stars in different mass bins.

This result might seem surprising: the knowledge of the masses of particular stars does not provide additional information on (the number) of unobserved ones[7]. It can be argued that, for example, an excess or deficit of the observed number of stars in a given mass range constrains the total number of stars from being compatible with sampling effects. However, such arguments are valid for IMF inferences (which IMF shape is more probable, given some observations?), i.e., the problem of obtaining the IMF.

In our case, *a given IMF is assumed and the observations are a random realization of it.* The particular observed set may be a highly improbable (but still possible) realization of the assumed IMF. Nevertheless, whatever its a priori probability of happening, *it has actually happened*, and thus a posteriori probabilities must be obtained by taking this fact into consideration. In addition, since stellar masses are random variables (cf. Paper I), the occurrence of having a star (or a set of stars) with a given particular mass has no impact on the individual masses of the remaining stars.

The mode of $\Phi_{\mathcal{N}}(\mathcal{N}|N_{\rm a})$, $\mathcal{N}^{\rm mode}$ is obtained by equating to zero its first derivative with respect to $\mathcal{N}$, which, for large $\mathcal{N}$ values[8] , yields

$$\mathcal{N}^{\rm mode} \approx \frac{\beta - N_{\rm a}}{\ln p(m < m_{\rm a})}, \quad (14)$$

where we used the Stirling approximation of factorial functions and a first-order Taylor approximation of logarithm functions valid for $\beta \neq 0$. In the case of a flat distribution with $\beta = 0$, the approximate mode of the distribution is obtained by solving

$$p(m < m_{\rm a}) = \left(1 - \frac{N_{\rm a}}{\mathcal{N}^{\rm mode}}\right), \quad (15)$$

which coincides with the estimation of the probability $\widetilde{p}(m < m_{\rm a})$ for known $N_{\rm a}$ and $\mathcal{N}$. This means that $N_{\rm a}/p(m|m \geq m_{\rm a})$ provides the mode $\mathcal{N}^{\rm mode}$ of $\Phi_{\mathcal{N}}(\mathcal{N}|N_{\rm a})$ assuming a flat $\Phi_{\mathcal{N}}(\mathcal{N})$ distribution. However, we know that the initial cluster mass function (ICMF, $\Phi_{\mathcal{M}}(\mathcal{M})$) is not flat (Lada & Lada 2003;

[4] Hence, it includes the $N_{\rm a}$ subset with known stellar masses.

[5] In this case, we are distributing the known $N_{\rm a}$ stars in k *intervals* and not using the particular m values of known stars. Such intervals can be arbitrary and must only obey the condition $\sum_{i=1}^{k} n_i = N_{\rm a}$. So the index i here refers to the interval, not to a particular stellar mass.

[6] Since $\mathcal{N}$ is a discrete quantity, their pdf directly provides the probability. In addition, the distribution can be also expressed as a binomial distribution with $A(p_i, n_i)\, N_{\rm a}! = p(m \geq m_{\rm a})^{N_{\rm a}} = (1 - p(m < m_{\rm a}))^{N_{\rm a}}$.

[7] However, we note that such information is still relevant for the computation of $\mathcal{M}$: the individual masses of stars more massive than $m_{\rm a}$ provide the amount of mass in the mass range, $M_{\rm a}$.

[8] In practical terms it implies large $N_{\rm a}$ values. Actually, $\Phi_{\mathcal{N}}(\mathcal{N}|N_{\rm a})$ is a discrete distribution, hence not derivable, but the formulae provide a reasonable value as far as the Stirling approximation of factorial functions are valid, i.e., $\mathcal{N}$, $N_{\rm a}$, and $\mathcal{N} - N_{\rm a}$ larger than 15.

name	$\mathcal{M}$	$M_{\rm a}$	$\mathcal{N}$	$N_{\rm a}$	$\langle \tilde{m}_{\rm a} \rangle$	$\log p_{\rm nor}(N_{\rm a}\|\mathcal{N})$
Tau.						
#1	10.6	7.6	20	8	0.95	-1.06
#2	15.5	11.5	30	12	0.96	-1.55
#3	8.1	5.9	19	8	0.74	-1.20
#4	22.7	21.7	24	18	1.20	-7.56
#5	8.2	8.0	14	10	0.80	-3.91
#6	17.7	14.7	31	14	1.05	-2.39
#7	16.1	13.9	24	13	1.07	-3.20
#8	12.3	10.2	16	5	2.05	-0.34
(field)	88.5	72.3	174	73	0.99	-10.15
ChaI						
#1	3.7	1.7	12	2	0.85	0.00
#2	40.5	25.9	96	20	1.30	-0.04
#3	21.7	16.4	43	16	1.03	-1.69
(field)	42.6	30.1	86	27	1.11	-1.58
Lup.3						
#1	18.2	13.4	36	11	1.22	-0.62
(field)	18.1	12.9	34	11	1.17	-0.76
IC348						
#1	111.9	87.6	186	65	1.35	-5.43
#2	3.1	0.5	11	1	0.53	-0.08
(field)	78.2	51.7	166	35	1.48	-0.08

Table 1. Data form stellar associations by Kirk & Myers (2011). We show the total mass ($\mathcal{M}$), the mass into stars more massive than $m_{\rm a}$ ($M_{\rm a}$), the total number of stars ($\mathcal{N}$), the number of stars more massive than $m_{\rm a}$ ($N_{\rm a}$), the estimation of the mean mass for stars more massive than $m_{\rm a}$ ($\langle \tilde{m}_{\rm a} \rangle = \langle \tilde{m} | m \geq m_{\rm a} \rangle$), and the logarithm of the probability that a cluster with $\mathcal{N}$ stars following the assumed IMF would have $N_{\rm a}$ stars with mass equal or larger than $m_{\rm a}$ divided by the maximum of such distribution ($\log p_{\rm nor}(N_{\rm a}|\mathcal{N})$). The $m_{\rm a}$ value is set to 0.5 $M_\odot$.

Piskunov et al. 2008) and that it must be somehow related to $\Phi_{\mathcal{N}}(\mathcal{N})$ (cf. Eq. 4), although we are not able to establish its functional form. Whatever equation we use to obtain $\mathcal{N}^{\rm mode}$, we are left in the uncomfortable situation of mixing a mean value ($\langle m | m < m_{\rm a} \rangle$) with a mode value $\mathcal{N}^{\rm mode}$ to obtain an inference about $\mathcal{M}$. However, we have no means to give a meaning of this inference: Is it a mean, a mode, on any other parameter?

This suggests that it is better to use the resulting probability distribution of $\mathcal{N}(N_{\rm a})$ and obtain the corresponding $\Phi_{\mathcal{M}}[\mathcal{M}|N_{\rm a}, \Phi_{\mathcal{N}}(\mathcal{N})]$ to make inferences about $\mathcal{M}$. In addition, this way to proceed is in agreement with the International Organization for Standardization (ISO), which recommends expressing the uncertainty in the results as a pdf[9].

4. Use cases

Having presented the probabilistic framework and the related information trade-off, we can compare the probabilistic methodology and the distribution by number methodology to obtain $\mathcal{M}$ and $\mathcal{N}$. For comparison purposes, we have used the data from Kirk & Myers (2011) to illustrate the differences. The data contain the observed masses for individual stars belonging to 14 young stellar groups in four different regions. They also contain the stellar mass of field stars in the four analyzed regions. Table 1 shows the identifier of the cluster along with the values of $\mathcal{M}$, $M_{\rm a}$, $\mathcal{N}$, $N_{\rm a}$, and the estimation of the mean mass, $\langle \tilde{m}_{\rm a} \rangle = \langle \tilde{m} | m \geq m_{\rm a} \rangle$ from the census of stars with $m \geq m_{\rm a}$. Kirk & Myers (2011) state that their mass estimates are valid with a relative error of 50%; in this work we assume that the tabulated values can be taken at face value without errors. They also state that their census is complete at a 90% level down to 0.08 $M_\odot$; hence their total mass estimation would be actually a lower limit of the real value. Whatever the case, we assume again that the $\mathcal{M}$ values obtained from the data can be use at face value without errors. Finally, we assume that the data is complete at 100% down to $m_{\rm a} = 0.5 M_\odot$. We use this $m_{\rm a}$ value to illustrate the $\mathcal{M}$ inference in scenarios 2, 3, and 4 in the introduction.

As reference, the IMF used here produces $\langle m \rangle = 0.46 M_\odot$, $\langle m | m \geq m_{\rm a} \rangle = 1.64 M_\odot$, and $p(m | m \geq m_{\rm a}) = 0.19$. We can make a first-order test about the compatibility of the cluster data with the assumed IMF by computing the probability of having a given $N_{\rm a}$ number of stars with mass larger than $m_{\rm a}$ in a cluster with $\mathcal{N}$ stars. It can be done by dividing the IMF into two bins, $[{\rm m_{low}}, m_{\rm a})$ and $[m_{\rm a}, {\rm m_{up}})$, and using the probability in each bin to define a binomial distribution. The logarithm value of the resulting probabilities normalized to the maximum value of the distribution, $\log p_{\rm nor}(N_{\rm a}|\mathcal{N})$, are shown in column 7 of Table 1[10]. In this test we see that our hypothesis about the validity of the used IMF in all the associations is actually questionable for the stars in Taurus field, Taurus #4, and IC348 #1, and would produce some problems in the analysis of Taurus #5, #7, and #6.

4.1. Distribution-by-number methodology

The distribution-by-number methodology considers that the IMF can be used with an arbitrary normalization. Such normalization can be either to $\mathcal{N}$ or $\mathcal{M}$, which implies multipling $\phi(m)$ by $\mathcal{N}$ or $\mathcal{M}/\langle m \rangle$, respectively. In addition, it is assumed that $\mathcal{N}$ and $\mathcal{M}$ are deterministically related by the relation

$$\mathcal{M} = \mathcal{N} \times \langle m \rangle \,. \tag{16}$$

This provides $\mathcal{M}$ in all the cases where $\mathcal{N}$ is given and vice versa. We can include additional information like $M_{\rm a}$ and $N_{\rm a}$ to make alternative inferences about $\mathcal{M}$. Following the procedure of this paper, the most information is included using a formula similar to Eq. 11:

$$\mathcal{M} = M_{\rm a} + (\mathcal{N} - N_{\rm a}) \langle m | m < m_{\rm a} \rangle \,. \tag{17}$$

However, we can choose to use only partial information, such as the contribution of $M_{\rm a}$ to the total budget. Then the ratio $\mathcal{M}/M_{\rm a}$ is constant, and is equal to the ratio of $m \times \phi(m)$ integrated in the whole range, $\langle m \rangle$, over the same function integrated in the $m_{\rm a}$, ${\rm m_{up}}$ range. As a result, $\mathcal{M}$ is:

$$\mathcal{M} = \frac{M_{\rm a} \times \langle m \rangle}{\int_{m_{\rm a}}^{\rm m_{up}} m\, \phi(m)\, {\rm d}m}. \tag{18}$$

On the other hand, we could choose to use the contribution of $N_{\rm a}$ to the total budget. Then the ratio $\mathcal{N}/N_{\rm a}$ is constant and is

[9] *Guide to the Expression of Uncertainty in Measurement* (International Organization for Standardization, Switzerland, 1995) `http://www.bipm.org/en/publications/guides/gum.html`.

[10] We note that a comparison of $\tilde{p}(m | m \geq m_{\rm a})$ and $p(m | m \geq m_{\rm a})$ does not produce a valid test about IMF compatibility, since the importance of the possible deviations depends on how many stars are in the sample (size of sample effects). Interestingly, IC348 #1, which deviates from the IMF in this test, is the system used as an example by Kirk & Myers (2011) to argue that their systems follows a Kroupa IMF (their Fig. 6). Although the shape of the IMF realization in IC348#1 would look like a Korupa IMF, the deviations (fluctuations) observed are actually too large compared with the expected ones taking into account the number of stars in the system.

equal to the ratio of $\phi(m)$ integrated in the whole range (that is, the unity) over the $\phi(m)$ integrated in the $m_{\rm a}$, $\rm m_{up}$ range. Since $\mathcal{M} = \mathcal{N} \times \langle m \rangle$, $\mathcal{M}$ is

$$\mathcal{M} = \frac{N_{\rm a} \times \langle m \rangle}{\int_{m_{\rm a}}^{\rm m_{up}} \phi(m)\,{\rm d}m}. \tag{19}$$

We could also choose to use just $M_{\rm a}$ and $N_{\rm a}$ values without the information about $\mathcal{N}$ (similar to Eq. 17 with some additional algebraic manipulation):

$$\mathcal{M} = M_{\rm a} + N_{\rm a}\ \langle m | m < m_{\rm a} \rangle\ \frac{\int_{\rm m_{low}}^{m_{\rm a}} \phi(m)\,{\rm d}m}{\int_{m_{\rm a}}^{\rm m_{up}} \phi(m)\,{\rm d}m}. \tag{20}$$

Equations 18, 19, and 20 produce an equal value of $\mathcal{M}$ as far as

$$\frac{M_{\rm a}}{N_{\rm a}} = \langle \tilde{m} | m \geq m_{\rm a} \rangle \rightarrow \langle m | m \geq m_{\rm a} \rangle\,,$$

and they will produce a result similar to Eq. 16 and 17 as far as, additionally,

$$\widetilde{p}(m | m \geq m_{\rm a}) = \frac{N_{\rm a}}{\mathcal{N}} \rightarrow p(m | m \geq m_{\rm a}).$$

In relation to the scenarios presented in the Introduction, scenario 2 (only $\mathcal{N}$ is observed) is described by Eq. 16. Scenario 3 ($\mathcal{N}$, $N_{\rm a}$, and $M_{\rm a}$ are known) can be described by Eqs. 16, 17, 18, 19, and 20, depending the information we choose to use, with Eq. 17 being the one that uses the most available information. Finally, scenario 4 can be described by Eqs. 18, 19, and 20, with Eq. 20 being the one that use the most available information.

The resulting $\mathcal{M}$ estimations from Kirk & Myers (2011) data employing this methodology are shown in Table 2, which uses different information from the cluster. The inferred $\mathcal{M}$ varies depending on the formulae (and hence the amount of not redundant information) used for the inference. The best result is obtained by Eq. 17, but unfortunately it does not have a practical application ($\mathcal{N}$ is unknown most of the times).

With respect to the equations that can be used in scenario 4 (the common observational case), Eq. 20 produce a value between the results of Eqs. 18 and 19. Also, since $\langle \tilde{m} | m \geq m_{\rm a} \rangle$ underestimates $\langle m | m \geq m_{\rm a} \rangle$ for the clusters in the given sample, Eq. 18 produces lower values than Eq. 19 (see Taurus #8 as the opposite example). The range of inferred $\mathcal{M}$ values covered by Eq. 18, 19 and 20 only include the observed $\mathcal{M}$ value in four cases (Taurus #8, Cha #1 and #2, and the field stars in IC348), suggesting a 20% rate of success (33% if we exclude the five clusters with possible strong deviations from the assumed IMF). In addition, we do not known which equation produces the more reasonable value (although Eq. 20 is preferred) nor do we have a possible evaluation accuracy associated to each case.

	$\mathcal{M}$ inferred					
	Sce.2	Sce.3	Sce.4			
name	Eq. 16	Eq. 17	Eq. 18	Eq. 19	Eq. 20	$\mathcal{M}$ obs.
Tau.						
#1	9.2	9.7	11.0	19.0	13.4	10.6
#2	13.8	14.7	16.7	28.6	20.3	15.5
#3	8.8	7.8	8.5	19.0	11.8	8.1
#4	11.1	22.7	31.3	42.8	34.9	22.7
#5	6.5	8.7	11.5	23.8	15.3	8.2
#6	14.3	17.7	21.2	33.3	24.9	17.7
#7	11.1	15.9	20.1	30.9	23.5	16.1
#8	7.4	12.2	14.8	11.9	13.9	12.3
field	80.3	90.1	104.5	173.7	125.8	88.5
ChaI						
#1	5.5	3.5	2.5	4.8	3.2	3.7
#2	44.3	39.3	37.5	47.6	40.6	40.5
#3	19.8	21.2	23.8	38.1	28.2	21.7
field	39.7	40.5	43.5	64.2	49.9	42.6
Lup.3						
#1	16.6	17.8	19.4	26.2	21.5	18.2
field	15.7	16.9	18.6	26.2	20.9	18.1
IC348						
#1	85.9	108.9	126.6	154.7	135.2	111.9
#2	5.1	2.3	0.8	2.4	1.3	3.1
field	76.6	74.8	74.7	83.3	77.3	78.2

Table 2. Inference of $\mathcal{M}$ employing the distribution-by-number methodology in the stellar associations by Kirk & Myers (2011), according different scenarios.

4.1.1. The probabilistic methodology

In the probabilistic case, pdfs are only used to describe unknown data, and observed data is used to define constraints over such unknown data, so that both types of data have different roles. In addition, the solution cannot be summarized in a single value, but as a distribution function. Although some summaries of such distribution (as the mean value) can be obtained analytically, such values do not necessarily have enough information, and the best method is to obtain the full distribution of possible solutions and work with it. We propose here the methodology to obtain the probability distribution of $\mathcal{M}$ when we know the individual masses of the most massive $N_{\rm a}$ stars, and we know that all stars equal to or more massive than $m_{\rm a}^{\rm obs}$ are included in the $N_{\rm a}$ set. The problem cannot be solved analytically since recursive convolutions involving power laws (such as the IMF) have no analytical solution. So we can only propose the following step-by-step procedure:

1. *Obtain the distribution of $\mathcal{N}$, $\Phi_{\mathcal{N}}(\mathcal{N}|N_{\rm a})$, which can be inferred from the data using Eq.13.* We stress again that an assumption about $\Phi_{\mathcal{N}}(\mathcal{N})$ is required. We note that the result would be quite dependent on the lower limit assumed in the $\Phi_{\mathcal{N}}(\mathcal{N})$ distribution.
2. *Compute the distribution of $\Phi_{M_{\rm not-obs}}(M_{\rm not-obs}|N_i)$ for the possible values of $N_i = \mathcal{N} - N_{\rm a}$ values obtained from the previous distribution*. The distribution provides the distribution of possible values of the total mass *from the unknown stars*, $M_{\rm not-obs}$, that is, $\mathcal{M}$ is actually constrained to the non-observed stellar masses $m < m_{\rm a}^{\rm obs}$, so we must use a constrained IMF to describe what we do not know, $\phi(m|m < m_{\rm a})$. Such $\Phi_{M_{\rm not-obs}}(M_{\rm not-obs}|N_i)$ distributions can be computed either by Monte Carlo simulations or by a numerical self-convolution.
3. *Compute the distribution of $\Phi_{\mathcal{M}}(\mathcal{M}|M_{\rm a}, N_{\rm a})$*. This is done by weighting the previous $\Phi_{M_{\rm not-obs}}(M_{\rm not-obs}|N_i)$ distributions by the probabilities of each N_i value provided by $\Phi_{\mathcal{N}}(\mathcal{N}|N_{\rm a})$ and including the contribution to the total mass of the observed stars.

We note that these two last steps can be done by means of Monte Carlo simulations, which sample the *discrete* distribution $\Phi_{\mathcal{N}}(\mathcal{N}|N_{\rm a})$ to obtain different N_i values, and by sampling the

	$\mathcal{M}$ inferred in scenario 2						
name	mean	mode	95.4% CL		68.3% CL		$\mathcal{M}$ obs
Tau.							
#1	9.2	5.9	2.7	21.2	3.7	9.7	10.6
#2	13.8	9.6	4.8	29.8	6.8	14.8	15.5
#3	8.8	5.2	2.5	20.5	3.5	9.0	8.1
#4	11.1	7.3	3.5	24.5	5.0	11.5	22.7
#5	6.5	3.7	1.5	15.5	2.5	7.0	8.2
#6	14.3	10.1	4.9	30.9	6.9	14.9	17.7
#7	11.1	7.3	3.5	24.5	5.0	11.5	16.1
#8	7.4	4.5	1.7	17.2	2.7	7.7	12.3
field	80.3	67.7	47.5	132.0	55.5	87.0	88.5
ChaI							
#1	5.5	2.9	1.2	13.7	1.7	5.7	3.7
#2	44.3	35.0	22.8	80.8	27.8	47.8	40.5
#3	19.8	13.6	7.8	40.8	10.3	20.8	21.7
field	39.7	31.0	19.8	73.8	24.3	42.8	42.7
Lup.3							
#1	16.6	11.8	6.0	35.0	8.5	18.0	18.2
field	15.7	10.7	6.0	33.5	8.0	17.0	18.2
IC348							
#1	85.8	71.0	51.2	140.2	60.2	93.2	111.9
#2	5.1	2.8	1.0	12.5	1.5	5.0	3.1
field	76.7	64.3	45.1	127.6	52.6	83.1	78.2

Table 3. Inference of $\mathcal{M}$ employing probabilistic methodology for the stellar associations by Kirk & Myers (2011) in scenario 2, using the observed value of $\mathcal{N}$.

	$\mathcal{M}$ inferred in scenario 3						
name	mean	mode	95.4% CL		68.3% CL		$\mathcal{M}$ obs
Tau.							
#1	9.7	9.9	8.6	10.6	9.1	10.1	10.6
#2	14.7	14.4	13.7	16.2	14.2	15.7	15.5
#3	7.8	7.6	6.8	8.8	7.3	8.3	8.1
#4	22.7	22.5	21.8	23.3	22.3	23.3	22.7
#5	8.7	8.8	8.0	9.5	8.0	9.0	8.2
#6	17.7	17.4	16.7	19.2	16.7	18.2	17.7
#7	15.9	15.6	14.9	16.9	15.4	16.4	16.1
#8	12.2	11.9	11.2	13.2	11.7	12.7	12.3
field	90.2	90.2	87.9	92.9	88.9	91.4	88.5
ChaI							
#1	3.5	3.4	2.6	4.6	3.1	4.1	3.7
#2	39.3	39.5	37.3	41.8	38.3	40.8	40.5
#3	21.2	21.0	19.7	22.7	20.2	21.7	21.7
field	40.5	40.6	38.4	42.4	39.4	41.4	42.7
Lup.3							
#1	17.8	17.5	16.3	19.3	17.3	18.8	18.2
field	16.9	16.9	15.7	18.2	16.2	17.7	18.2
IC348							
#1	109	109	106	112	108	111	112
#2	2.3	2.2	1.4	3.4	1.9	2.9	3.1
field	74.8	74.7	71.9	77.9	73.4	76.4	78.2

Table 4. Inference of $\mathcal{M}$ employing probabilistic methodology for the stellar associations by Kirk & Myers (2011) in scenario 3, using the observed value of $\mathcal{N}$, N_a, M_a and $m_a = 0.5\mathrm{M}_\odot$.

constrained IMF with this number of stars. The previous procedure covers scenarios 2 and 3 by applying only step 2: obtain $\Phi_{\mathcal{M}}(\mathcal{M}|\mathcal{N})$ or $\Phi_{\mathcal{M}}(\mathcal{M}_{\mathrm{not-obs}}|N_i)$ for a known $\mathcal{N}$.

We applied this methodology to the set of clusters of Kirk & Myers (2011) under different scenarios by means of Monte Carlo simulations. The distribution of solutions for each cluster in each scenario was sampled by 10^7 Monte Carlo simulations, and the resulting distribution was binned in intervals with $\Delta\mathcal{M} = 0.5\mathrm{M}_\odot$. We note that in scenario 4 the simulations sample both the IMF and the assumed $\Phi_{\mathcal{N}}(\mathcal{N})$ distributions (power laws with $\beta = 0$ and $\beta = 2$). Therefore, the simulations span a larger $\mathcal{M}$ range and an additional uncertainty is expected for the confidence interval estimations.

Table 3 shows the resulting mean, mode, and 68.3% (equivalent to 1σ in a Gaussian distribution) and 95.4% (equivalent to 2σ in a Gaussian distribution) confidence intervals around the mode for scenario 2. As expected, the mean value of the distribution coincides with the result of Eq.16 shown in Table 2. All observed $\mathcal{M}$ are in the 94.5% confidence interval around the mode, although only 27% are in the 68.3% confidence interval, being the observed $\mathcal{M}$ larger than the range quoted in such interval.

Table 4 shows the results of the $\mathcal{M}$ distribution for scenario 3, which includes a larger amount of information. The mean and mode of the distribution coincides (hence the distribution is symmetric), and the mean value is also coincident to the result of Eq. 17, as expected. However, in this case we can evaluate how good this estimation actually is (and hence the distribution by number estimation). Taking favorable round-around cases, 17% of the clusters (i.e., field stars in ChaI, IC348 #1, and field stars in IC348) are outside the 2σ range, 83% are in the 2σ range, and 67% are in the 1σ range (i.e., 12 clusters). Given the low number of clusters for this study, we find this result partially consistent with a standard methodology. However, in theory, we would expect only one cluster outside the 2σ range, although we can invoke the use of a low number of clusters for this study.

	$\mathcal{M}$ inferred in scenario 4 with $\Phi_{\mathcal{N}}(\mathcal{N})$ = cte						
name	mean	mode	95.4% CL		68.3% CL		$\mathcal{M}$ obs
Tau.							
#1	14.2	13.4	9.7	19.7	11.2	16.2	10.6
#2	21.1	20.1	15.4	27.4	17.4	23.4	15.5
#3	12.5	11.8	8.0	18.0	9.5	14.5	8.1
#4	35.6	34.8	28.5	43.5	31.5	39.0	22.7
#5	16.0	15.0	10.8	21.8	12.8	18.3	8.2
#6	25.7	24.7	19.4	32.4	21.9	28.4	17.7
#7	24.2	23.1	18.4	30.9	20.4	26.9	16.1
#8	14.6	14.0	10.7	18.7	12.2	16.2	12.3
field	127	126	112	142	119	134	89
ChaI							
#1	3.9	3.0	1.7	7.2	1.7	4.7	3.7
#2	41.3	40.3	34.1	49.6	36.6	44.6	40.5
#3	28.9	28.4	22.2	36.2	24.7	31.7	21.7
field	50.6	49.5	41.7	59.7	45.7	54.7	42.7
Lup.3							
#1	22.2	21.4	16.7	28.2	18.7	24.7	18.2
field	21.7	20.9	16.1	27.6	18.1	24.1	18.2
IC348							
#1	136	135	122	150	128	142	112
#2	2.0	0.8	0.5	4.5	0.5	2.5	3.1
field	78.0	76.9	68.2	88.7	72.2	82.7	78.2

Table 5. Inference of $\mathcal{M}$ employing probabilistic methodology for the stellar associations by Kirk & Myers (2011), using the value of N_a, M_a and $m_a = 0.5\mathrm{M}_\odot$ and assuming a flat $\Phi_{\mathcal{N}}(\mathcal{N})$ distribution.

An additional outcome of this study is that, although Eq. 17 produces results similar to the observations, it does not necessarily provide a fully compatible (e.g., at 1σ level) result. Again, this enforces the idea of using the whole pdf of possible solutions instead a summary (like the confidence interval range) of it.

Tables 5 and 6 show the results of applying this methodology using flat and power law $\Phi_{\mathcal{N}}(\mathcal{N})$ distributions ($\beta = 0$ and $\beta = 2$, respectively) in the range from $\mathcal{N} = N_a$ to $\mathcal{N} = 4000$ stars. The

name	$\mathcal{M}$ inferred in scenario 4 with $\Phi_{\mathcal{N}}(\mathcal{N}) \propto \mathcal{N}^{-2}$. mean	mode	95.4% CL		68.3% CL		$\mathcal{M}$ obs
Tau.							
#1	12.7	11.8	8.6	17.6	10.1	14.6	10.6
#2	19.6	18.8	14.5	25.5	16.0	21.5	15.5
#3	11.1	10.2	6.9	15.9	8.4	12.9	8.1
#4	34.1	33.4	27.7	41.7	30.2	37.2	22.7
#5	14.6	13.7	10.0	20.0	11.5	16.5	8.2
#6	24.2	23.5	18.2	30.7	20.7	26.7	17.7
#7	22.7	22.1	17.4	28.9	19.4	25.4	16.1
#8	13.2	12.5	10.2	16.7	10.7	14.2	12.3
field	125	124	111	140	118	132	89
ChaI							
#1	2.6	2.0	1.7	4.7	1.7	3.2	3.7
#2	39.9	39.0	32.7	47.7	35.7	43.2	40.5
#3	27.5	26.6	21.3	34.3	23.8	30.3	21.7
field	49.2	48.2	41.0	58.5	44.0	53.0	42.7
Lup.3							
#1	20.7	19.9	15.7	26.2	17.7	23.2	18.2
field	20.2	19.4	15.2	25.7	17.2	22.7	18.2
IC348							
#1	135	133	121	149	127	141	112
#2	0.8	0.8	0.5	2.0	0.5	1.0	3.1
field	76.6	76.0	66.7	86.7	71.2	81.2	78.2

Table 6. Inference of $\mathcal{M}$ employing probabilistic methodology for the stellar associations by Kirk & Myers (2011), using the value of N_a, M_a and $m_a = 0.5\mathrm{M}_\odot$ and assuming a power-law $\Phi_{\mathcal{N}}(\mathcal{N})$ distribution with $\beta = 2$.

first result is that mean and mode values of the distribution are not equal in general, and the distribution is not symmetric, but j-shaped. The mode in the case of a flat $\Phi_{\mathcal{N}}(\mathcal{N})$ distribution is similar to the result obtained by Eq. 20. In this case, the observed $\mathcal{M}$ of seven clusters are outside the 2σ confidence interval (actually, the clusters with lower $p_{\mathrm{nor}}(N_a|\mathcal{N})$ value quoted before and ChaI#3). If we neglect the six clusters with the larger deviations from the IMF, we obtain a result showing that 9% of the cluster are outside the 2σ interval, 91% of the cluster are in the 2σ interval, and 55% of the clusters are in the 1σ interval. This is a reasonable result of any statistical test.

Finally, the results of Eqs. 18 and 20 are within the 2σ range in the case of a flat $\Phi_{\mathcal{N}}(\mathcal{N})$ distribution, but the results of Eq. 19 (estimation from the extrapolations of the observed N_a) produce larger values than the upper limit of 2σ.

5. Discussion

We have shown in this work that the determination of cluster masses is not so trivial as supposed in the literature. The distribution-by-number methodology uses known data to *determine* unknown data, whereas the probabilistic methodology uses known data to *constrain* unknown data. The problem is also related to the trade-off between unknown data and probability. When we use a pdf, like the IMF, to make inferences about unknown data, we implicitly renounce obtaining actual values of the inferred quantity. The price is to renounce precision in favor of accuracy. In contrast, the distribution-by-number methodology favors precision and renounces accuracy. The difference is in the algebra (and the logic reasoning) used in each of the methodologies to manipulate formulae. The distribution-by-number methodology uses standard algebra, where symbols are just mathematical expressions without added meaning. The probabilistic methodology follows the algebra of probability, which implies a clear identification of the known and the (random) variables we aim to describe by a probability distribution.

As an example, the equation

$$\mathcal{N} \times p(m \geq m_a^{\mathrm{obs}}) = N_a$$

provides an *estimation* of the number of stars with mass equal or larger than m_a^{obs} in a cluster with $\mathcal{N}$ stars. But such an estimation is not necessarily a mean value nor a mode value (cf. Paper I for the case that $N_a = 1$). In that case, we know $\mathcal{N}$; hence, we are working with a $\Phi_{N_a}(N_a|\mathcal{N})$ distribution. The inversion of the equation, that is,

$$\mathcal{N} = \frac{p(m \geq m_a^{\mathrm{obs}})}{N_a},$$

provides the modal value $\mathcal{N}^{\mathrm{mode}}$ of the distribution $\Phi_{\mathcal{N}}(\mathcal{N}|N_a)$ when a flat distribution of $\mathcal{N}$ values is assumed, (i.e., $\Phi_{\mathcal{N}}(\mathcal{N})$ = constant). The distribution $\Phi_{\mathcal{N}}(\mathcal{N})$ appears naturally when the Bayes' theorem is used. This is a natural result when we realize that, since $\mathcal{N}$ is unknown, we need its probability distribution to make inferences about it, and that the "innocent" algebraic manipulation we have done has a completely different meaning than the one we would expect.

5.1. To $\Phi_{\mathcal{N}}(\mathcal{N})$ or not to $\Phi_{\mathcal{N}}(\mathcal{N})$?

We are now in the uncomfortable situation of having to assume a $\Phi_{\mathcal{N}}(\mathcal{N})$ distribution in the inference of $\mathcal{N}$ and $\mathcal{M}$. However, the relevance of the $\Phi_{\mathcal{N}}(\mathcal{N})$ in the inference of $\mathcal{M}$ is also dependent on the value of m_a and N_a. In a back-of-the-envelope argument, the effect of a power-law $\Phi_{\mathcal{N}}(\mathcal{N})$ distribution is to decrease N_a by β stars (cf. Eq. 14 used for $\mathcal{N}^{\mathrm{mode}}$ estimation). Hence, the larger N_a, the lower the dependence of the $\mathcal{M}$ estimation on $\Phi_{\mathcal{N}}(\mathcal{N})$. Of course, the way to increase N_a is to be complete down to the lowest m_a possible.

In the cases where the $\mathcal{M}$ inference strongly depends on our choice of $\Phi_{\mathcal{N}}(\mathcal{N})$, we must be guided by our knowledge of the physical system environment and the scientific goal of the analysis. A flat $\Phi_{\mathcal{N}}(\mathcal{N})$ assumes that there is no previous knowledge about the system *environment*, so it looks like good option in the case of isolated systems and when we are only interested in the system properties.

However, the situation varies if we are interested in a cluster that *we know* is in a supercluster environment or is the result of molecular cloud fragmentation. In these cases, depending on our knowledge and hypothesis about star formation (SF), we can consider that such fragmentation is the result of a high-order structure; hence, the particular cluster is not an isolated entity. This would imply that some values of $\mathcal{N}$ or $\mathcal{M}$ are more probable than others, and this information must be taken into account in the inference of $\mathcal{N}$ and $\mathcal{M}$ of the particular cluster.

We must stress here that the proposed method only applies to $\Phi_{\mathcal{N}}(\mathcal{N})$ distributions, and not to $\Phi_{\mathcal{M}}(\mathcal{M})$ ones. The case of $\Phi_{\mathcal{N}}(\mathcal{N})$ is easily implemented as far as it is related to sampling theory and the number of the elements in the sample is the relevant quantity. The inclusion of $\Phi_{\mathcal{M}}(\mathcal{M})$ is not so trivial, since it depends implicitly on a $\Phi_{\mathcal{N}}(\mathcal{N})$ distribution. However, such distribution can not be obtained analytically (the convolution problem is not analytic in general cases). In addition, since $\Phi_{\mathcal{N}}(\mathcal{N})$ is a discrete distribution, we have a large, but finite (and hence computable), number of cases. This is not true for $\Phi_{\mathcal{M}}(\mathcal{M})$ because it is a continuous function and the possible solutions that

a combination of $\mathcal{N}$ stars produces a particular $\mathcal{M}$ is infinite. At this moment, the only solution is to use $\Phi_{\mathcal{M}}(\mathcal{M})$ as a proxy for $\Phi_{\mathcal{N}}(\mathcal{N})$, which would be valid for situations where we know a priori that the minimum possible number of stars is large (i.e., $N_{\rm a}$ is large, or we have additional information about a minimum number of stars in the cluster).

Finally, the situation also changes if we are interested in obtaining $\Phi_{\mathcal{N}}(\mathcal{N})$ or $\Phi_{\mathcal{M}}(\mathcal{M})$ from a set of clusters. Following Tarantola (2006), the most viable way is to make an iterative process. First, assume a $\Phi_{\mathcal{N},0}(\mathcal{N})$ distribution and compute resulting distributions of $\mathcal{N}_i$ and $\mathcal{M}_i$ for each cluster. After that, combine such distributions to obtain *from the sample* the global distributions $\Phi_{\mathcal{N},1}(\mathcal{N})$ and $\Phi_{\mathcal{M},1}(\mathcal{M})$. If $\Phi_{\mathcal{N},1}(\mathcal{N}) \neq \Phi_{\mathcal{N},0}(\mathcal{N})$, then $\Phi_{\mathcal{N},0}(\mathcal{N})$ is not a self-consistent hypothesis. However, we must be aware that this does not prove that $\Phi_{\mathcal{N},1}(\mathcal{N})$ and $\Phi_{\mathcal{M},1}(\mathcal{M})$ are self-consistent hypotheses! The only way to achieve a self-consistent hypothesis is iterate the process until $\Phi_{\mathcal{N},j-1}(\mathcal{N}) = \Phi_{\mathcal{N},j}(\mathcal{N})$ being the $j-1$ distribution is the one used as input and the j distribution is the resulting one, along with testing if the resulting $\Phi_{\mathcal{M},j}(\mathcal{M})$ distributions also obey such a condition (a cross validation). However, we stress again that such a cross-validation process is a requirement that depends on the $N_{\rm a}$ value and that for large enough $N_{\rm a}$ values, the resulting $\Phi_{\mathcal{M}}(\mathcal{M}|M_{\rm a}, N_{\rm a})$ solution for the $\mathcal{M}$ distribution of a cluster is almost $\Phi_{\mathcal{N}}(\mathcal{N})$ independent.

6. Conclusions

Throughout this work, we have explicitly developed the use of the IMF to obtain different physical parameters of stellar systems from limited information. We made extensive use of the IMF as a pdf, which allowed us to make proper use of probability theory and, in particular, the properties of sampling distributions (where the total number of stars in the system is included) and conditional probabilities.

We studied the methodology to obtain the distribution of possible $\mathcal{N}$ and $\mathcal{M}$ values from the knowledge of the set of the most massive stars in the system. The result is dependent on the values of $m_{\rm a}$ and $N_{\rm a}$, and on the hypothesis about the overall distribution of the number of stars in clusters $\Phi_{\mathcal{N}}(\mathcal{N})$, including the limits of such distribution (especially the lower one).

Acknowledgements. We acknowledge Kevin Covey for discussions of the similarities and differences of $\Phi_{\mathcal{N}}(\mathcal{N})$ and $\Phi_{\mathcal{M}}(\mathcal{M})$ and their implications in the modeling of clusters and galaxies, which have been very useful for this paper and related works. We acknowledge the suggestions of the referee, Peter Anders, which greatly improved the clarity of the paper. This work has been supported by the MICINN (Spain) through the grants AYA2007-64712, AYA2010-15081, AYA2010-15196, AYA201122614, AYA2011-29754-C03-01, AYA2008-06423-C03-01/ESP, AYA2010-17631, programs UNAM-DGAPA-PAPIIT IA101812 and CONACYT 152160, Mexico, and co-funded under the Marie Curie Actions of the European Commission (FP7-COFUND)

References

Bayo, A., Barrado, D., Stauffer, J., et al. 2011, A&A, 536, A63
Briceño, C., Luhman, K. L., Hartmann, L., Stauffer, J. R., & Kirkpatrick, J. D. 2002, ApJ, 580, 317
Bouvier, J., Stauffer, J. R., Martin, E. L., et al. 1998, A&A, 336, 490
Cerviño, M., & Luridiana, V. 2006, A&A, 451, 475
Cerviño, M., Román-Zuñiga, C. Luridiana, V. et al. A&A, (submited, Paper I)
Kirk, H., & Myers, P. C. 2011, ApJ, 727, 64
Kroupa, P. 2001, MNRAS, 322, 231
Kroupa, P. 2002, Science, 295, 82
Lada, C. J., & Lada, E. A. 2003, ARA&A, 41, 57
Luhman, K. L., Stauffer, J. R., Muench, A. A., et al. 2003, ApJ, 593, 1093
Oliveira, J. M., Jeffries, R. D., & van Loon, J. T. 2009, MNRAS, 392, 1034
Piskunov, A. E., Kharchenko, N. V., Schilbach, E., et al. 2008, A&A, 487, 557
Scalo 1986 Scalo, J. M. 1986, *Fun. Cosm. Physics* 11, 1.
Selman, F. J., & Melnick, J. 2008, ApJ, 689, 816
Tarantola, A. 2006, Nature Physics 2, 492
Weidner, C., & Kroupa, P. 2006, MNRAS, 365, 1333